AMERICAN JOURNAL OF PHYSICAL ANTHROPOLOGY 111:463–478 (2000)

Statistical Power Comparisons Among Alternative Morphometric Methods

F. JAMES ROHLF*
Department of Ecology and Evolution, State University of New York, Stony Brook, NY 11794-5245

KEY WORDS power surfaces; Kendall shape space; tangent space; shape coordinates; EDMA; inter-landmark distances; multivariate analysis; morphometrics

ABSTRACT This paper compares the statistical power of various tests that have been proposed to test for equality of shape in two populations. Power surfaces are computed with emphasis on the simplest case of three points in the plane (i.e., landmarks at the vertices of a triangle). Goodall's ([1991] J Roy Stat Soc Serb 53:285–339) F-test was found to have the highest power followed by T^2-tests using Kendall tangent space coordinates. Power for T^2-tests using Bookstein shape coordinates was good if the baseline was not the shortest side of the triangle. The Rao and Suryawanshi ([1996] Proc Natl Acad Sci 93:12132–12136 and [1998] Proc Natl Acad Sci 95:4121–4125) shape variables had much lower power when triangles were not close to being equilateral. Power surfaces for the EDMA-I T statistic revealed very low power for many shape comparisons including those between very different shapes. Power surface for the EDMA-II Z statistic were also complicated and depended strongly on the choice of baseline used for size scaling. The type I error rate was also often not correct for this method. Results for more than three landmarks are also presented. The implications of the results for practical applications of morphometrics are discussed. Am J Phys Anthropol 111:463–478, 2000.

The fundamental advances of geometric morphometrics over traditional approaches (multivariate morphometrics, e.g., Reyment et al., 1984) are in the development of powerful statistical methods based on models for shape variation of entire configurations of points corresponding to the locations of morphological landmarks. This is in contrast to the application of standard multivariate methods to ad hoc collections of distances, angles, and ratios based on these same landmarks. The shape of a set of p morphological landmark points is defined as those attributes of the configuration of points that are invariant to the effects of translation, rotation, and scale. The positions of the points can be captured conveniently by their coordinates (x, y for images or x, y, z for actual organisms). There are several ways in which the coordinates can be adjusted so that they are unaffected by variation in a specimen's location, orientation, and scale. There has been some controversy over these alternative approaches. This paper compares several approaches in terms of their statistical power in tests for shape differences in two samples. While a part of geometric morphometrics, the analysis of the shapes of outlines will not be considered here. Rohlf and Marcus (1993)

Grant sponsor: National Science Foundation; Grant number: 1BN-9728160.

*Correspondence to: F. James Rohlf, Department of Ecology and Evolution, State University of New York, Stony Brook, NY 11794-5245. E-mail rohlf@life.bio.sunysb.edu

Received 11 March 1999; accepted 24 October 1999.

TABLE 1. Tests for shape differences

Code	Statistic	Shape variables	Reflections	Reference
GoodF	Goodall *F*-test	Procrustes aligned	Kept separate	Goodall (1991)
KendTS	T^2-test	Kendall tangent space or partial warps + uniform component	Kept separate	Rohlf (1995)
BookSC	T^2-test	Bookstein shape coordinates	Kept separate	Bookstein (1991)
Rao-d	T^2-test	Linear combinations of log size-scaled interlandmark distances	Not distinguished	Rao and Suryawanshi (1996)
Rao-a	T^2-test	Subset of interior angles	Not distinguished	Rao and Suryawanshi (1998)
EDMA-I	*T*, nonparametric bootstrap	Interlandmark distances	Not distinguished	Lele and Richtsmeier (1991)
EDMA-II	*Z*, parametric bootstrap	Size-scaled inter-landmark distances	Not distinguished	Lele and Cole (1996)

give a general overview of the field of geometric morphometrics. Bookstein (1991) gives a comprehensive account while Small (1996) covers mathematical details. Dryden and Mardia (1998) cover many aspects of shape statistics. Marcus et al. (1996) includes both introductory material and examples of applications to many fields of biology and medicine.

Two rather different approaches for the statistical analysis of shape have been proposed:

1. Most analyses can be interpreted in terms of differences between configurations of landmarks that have been optimally superimposed using a least-squares procedure. These methods are based either on Procrustes distance as a metric to define a non-Euclidean shape space for comparing shapes or on Euclidean distance in a linear approximation to shape space (called a tangent space). Procrustes distance can be computed as the square root of the sum of squared differences between two optimally superimposed shapes. There has been a considerable amount of work in this area using Goodall's (1991) perturbation model for shape variation (see the next section). Kendall (1981, 1984, 1985) represents important initial work. The exact statistical distribution has even been worked out for certain cases (see Dryden and Mardia, 1998, for a discussion). The well-known thin-plate spline methods (Bookstein, 1991) are examples of methods that use the tangent space approximation. See Rohlf (1999b) for additional discussion about shape spaces and the relationships among these methods.
2. There are also several methods based on interlandmark distances. An example is Euclidean distance matrix analysis (EDMA) proposed by Lele and Richtsmeier (1991) and further elaborated in Lele (1993), Richtsmeier and Lele (1993), and Lele and Cole (1996). Statistical tests based on bootstrap procedures are used with these methods since analytical results are not available. Rao and Suryawanshi (1996) and Richtsmeier et al. (1998) proposed a related approach using standard multivariate methods on logs of size-scaled inter-landmark distances. Recently, Rao and Suryawanshi (1998) suggested the use standard multivariate methods on angles from triangulations of triplets of landmarks.

The purpose of the present study is to contrast the proposed methods in terms of their statistical power in testing for differences in mean shape between samples drawn from two different populations. Table 1 lists the methods that will be considered here and gives the codes that will be used to identify them. This study was done to assist users in deciding which of the competing methods should be used in practical applications. One prefers to use the method that has the highest statistical power under a set of statistical assumptions that are appropriate for the data. Rao and Suryawanshi

(1996) state that there is no unique way of choosing among alternative shape functions and that inference based on a particular choice of functions will be consistent with that based on another choice provided the probability distribution can be accurately specified. However, the statistical distributions are not known for many of the methods. It will be shown below that the proposed methods do not have the same statistical power for detecting differences in shape. In many cases, the differences in power are *very* large. Thus the various approaches are not equivalent in practice and they may lead to very different conclusions in practical applications. Power and efficiency of different morphometric methods (especially the BookSC, KendTS, EDMA-I, and EDMA-II methods) received a considerable amount of attention in the morphmet Internet discussion group in 1995 and 1996. Neither side seemed able to convince the other. It is hoped that the present contribution will resolve some of the issues.

The results presented below are largely based on sampling experiments using Monte Carlo simulations as in the studies by Coward and Conathy (1996) and by Lele and Cole (1996). However, those studies only investigated a few special cases so that it is difficult to draw general conclusions from their results. In the present study, the space of all possible differences between configurations of three landmarks is sampled (but we do have to resort to numerical examples for cases with more than just three landmarks). The statistical model used to model shape variation is presented in the next section. A limitation of the present study is that only the simplest case of homogeneous, independent, spherical variation at each landmark is considered. The sampling experiments emphasize the case of just $p = 3$ landmarks in the plane (triangular shapes) because it is possible to show the results as power surfaces in three-dimensional perspective plots. These power surfaces allow one to see how power varies when comparing the shape of one population against that of all possible alternative populations. This is important since it makes it possible to have a more intuitive understanding of the properties of the methods.

A STATISTICAL MODEL FOR SHAPE VARIATION

Goodall (1991) described a perturbation model as a simple model for variation in the positions of the landmarks around their mean locations. This model also allows for covariation between the landmarks. In this model the $p \times k$ matrix of coordinates for the p k-dimensional landmarks for the ith specimen is given by

$$\mathbf{X}_i = \alpha_i(\mu + \mathbf{E}_i)\Omega_i + \mathbf{1}\omega_i^t \qquad (1)$$

where α_i is a scale factor (size of the ith specimen), μ is the mean shape, $\mathbf{E}_i$ is a matrix of random errors (normally distributed with mean of zero), Ω_i is a $k \times k$ matrix describing the orientation of the ith specimen (reflections excluded), $\mathbf{1}$ is a vector of all ones, and ω_i is a vector specifying the location of the specimen in the digitizing plane (or solid). Parameters α_i, Ω_i, and ω_i are called nuisance parameters because they encode information unrelated to shape variation. The estimates of shape variation must be independent of these parameters. Matrix $\mathbf{E}_i$ (when strung out as a single column vector with kp elements) has a covariance matrix $\Sigma = \Sigma_p \otimes \Sigma_k$, where Σ_p is the covariance matrix for the landmark points (the rows of $\mathbf{E}_i$) and Σ_k is the covariance matrix for the dimensions (the columns of $\mathbf{E}_i$). The symbol $\otimes$ denotes a matrix direct product (also called a Kronecker product) of two matrices. In this study, only the simplest case was investigated—identical independent variation around each mean landmark position ($S_p = \sigma^2\mathbf{I}_p$ and $S_k = \mathbf{I}_k$). This is the type of variation one might expect from digitizing error.

Kendall (1981, 1984) showed that the shape space corresponding to the Procrustes metric is a $kp - k - k(k-1)/2$ one-dimensional manifold ($2p - 4$ dimensions for 2D and $3p - 7$ dimensions for 3D). A manifold is a generalization to higher dimensions of a curved surface in three dimensions. Small (1996) gives an introduction to shape spaces and discusses applications to morphometrics. For $p = 3$ and $k = 2$ (triangles in the plane), the manifold corresponds to the sur-

face of a sphere with radius = $\frac{1}{2}$ (a two-dimensional curved space). Every possible shape maps to a unique position in Kendall shape space. Under the perturbation model with digitizing-like error, one expects a circular scatter of points around the point corresponding to the mean shape.

In practical applications, a generalized least-squares Procrustes analysis (GLS, but also called a generalized Procrustes analysis, GPA) is performed to estimate a mean shape and to align the specimens to it. These aligned specimens are then used for the computation of their tangent space projections and partial warp scores that are used for statistical analyses. The space to visualize aligned specimens is the surface of a hyper-hemisphere (a hemisphere if $p = 3$) with the mean shape corresponding to its pole. This space has a radius of 1. Kendall tangent space represents an orthogonal projection of this space onto a plane tangent to its pole. For triangles this tangent space is a unit disk. See Rohlf (1999b) for more information about these relationships. Figure 1 illustrates the Kendall tangent space for triangles with an equilateral triangle arbitrarily used as the reference. Examples of triangles are shown for different locations in tangent space (different positions along the u_1, u_2 axes). For the perturbation model discussed above, projection of specimens into the Kendall tangent space also yields hyperspherical scatter as long as the point of tangency is close to the mean (as it should be in any practical application). Statistical power of tests for differences in shapes of triangles will be plotted as surfaces lofted above this space.

A stereographic projection of a distribution with hyperspherical density contours in Kendall shape space yields a distribution in tangent space with hyperspherical density contours since a stereographic projection is a conformal mapping (a transformation that preserves angles). However, unless the point of tangency is at the mean, the density contours will not be centered on the same point (i.e., the distribution will be skewed away from the origin). Bookstein shape coordinates are a special case corresponding to using as a reference a collinear triangle with the free point at the origin and a baseline defined by the other two landmarks. The expected variance of the scatter increases greatly for mean shapes further from the point of tangency—proportional to $(1 + x^2)^2$, where x is the distance in stereographic shape coordinates of the mean shape from the reference. For this reason the shortest edge should not be used as the baseline.

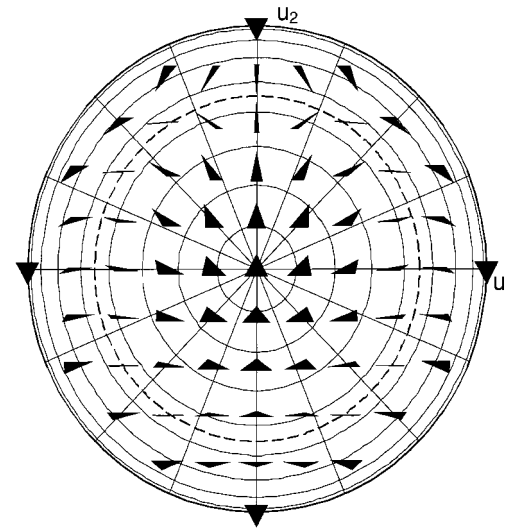

Fig. 1. Kendall tangent space for triangles using an eqilateral triangle as the reference shape. The u_1,u_2 axes correspond to Bookstein's linear Procrustes estimate of the uniform component. Each triangle is centered on its u_1,u_2 coordinates. The dashed circle (at a radius of $\sqrt{1/2}$ and a Procrustes distance of $\rho = \pi/4$ from the reference) is the locus of collinear triangles. Points outside of this circle correspond to triangles that are reflections of those within. All points on the outer circle (at a radius of 1 and a Procrustes distance of $\rho = \pi/2$ from the reference) correspond to the same triangle—the reflection of the equilateral triangle used as the reference. Exemplar triangles drawn using the AB edge as a baseline and landmark C as the free vertex. This figure was prepared using the tpsTri software.

Shape spaces corresponding to the other morphometric methods are very different. Rohlf (1999a) illustrates these shape spaces for triangles and shows the distribution for random samples of triangles for the shape variables defined by Rao and Suryawanshi (1996), Rao and Suryawanshi (1998), the EDMA-I statistic T, and the EDMA-II statistic Z. For these methods, the size and orientation of the scatter of points in their shape spaces are strongly influenced by the location of the mean in shape space. Since

little theoretical work seems to have been done in this area, the tpsTri software (Rohlf, 1999c) is useful for visualizing the statistical distribution of triangles using various morphometric methods.

STATISTICS TO TEST FOR DIFFERENCES IN MEAN SHAPE

Several procedures have been proposed to test for shape differences in samples from two populations. Table 1 lists those that will be investigated in this paper and gives the codes that will be used to identify them in this paper. A summary of the methods is given below.

Goodall's (1991) F-test

This test compares the Procrustes distance between the means of two samples to the amount of variation found within the samples. It uses a generalized least-squares Procrustes analysis (GLS, but also called a generalized Procrustes analysis, GPA) to compute the average shape for each sample. Goodall (1991) proposed the use of the following statistic to test for differences in mean shape for two samples of sizes n_1 and n_2.

$$F = \frac{d_F^2(\bar{X}_1, \bar{X}_2)}{\sum^{n_1} d_F^2(X_1, \bar{X}_1) + \sum^{n_2} d_F^2(X_2, \bar{X}_2)} \cdot \frac{n_1 + n_2 - 2}{(1/n_1) + (1/n_2)}, \quad (2)$$

where $d_F^2(\bar{X}_1, \bar{X}_2)$ is the squared full Procrustes distance between the mean shapes of the two samples

$$\sum^{n_1} d_F^2(X_1, \bar{X}_1) \quad \text{and} \quad \sum^{n_2} d_F^2(X_2, \bar{X}_2)$$

are the sums of squared full Procrustes distances between the fitted specimens in each sample and their mean shapes. For the full Procrustes distance one shape is allowed shrink in size so as to minimize the distance between the two shapes. Under the null hypothesis F will follow the F-distribution with m over $(n_1 + n_2 - 2)m$ degrees of freedom (m is the dimensionality of shape space, $2p - 4$ for 2D data or $3p - 7$ for 3D). A related approach, not investigated here, is to test Goodall's statistic using a permutation test (Bookstein, 1998).

In addition to assuming independent samples and that $\mathbf{E}_i$ follows the multivariate normal distribution, this test makes the rather restrictive assumptions that $\Sigma_p = \boldsymbol{\sigma}^2\mathbf{I}$ and $\Sigma_k = \mathbf{I}$ for both samples. Under the assumed model, this test is not valid (i.e., type I error rate becomes too large) when σ^2 is not small. Since d_F, d_P, ρ, and even Euclidean distance in Kendall tangent space are so similar, approximate tests can also be made substituting these other Procrustes distances. When the sphericity assumption is true (such as in the simulations in the present study), this test is able to exploit this fact and estimate many fewer parameters than are required for the usual T^2-test. This should result in higher power than for the usual T^2-tests—especially when sample sizes are small.

T^2 test using Kendall tangent space coordinates

This test is based on using GLS superimposition to compute the average shape for the entire dataset. Each specimen is then fit to this overall mean. Rohlf (1999b) shows that triangles corresponding to these aligned specimens lie on the surface of a unit hemisphere of the same dimensionality as Kendall's shape space. When shape variation is small, the distribution of points on this hemisphere can be satisfactorily approximated by an orthogonal projection onto an Euclidean tangent plane (if the overall mean is used as the point of tangency). Kent (1994) calls this space Kendall tangent space but Dryden and Mardia (1992) call it Kent tangent space (see Fig. 1 for an illustration for triangles). Partial warps (including Bookstein's 1996 linearized estimate of the uniform component) provide a convenient coordinate system for this space with the correct number of dimensions. It is also possible to work directly with the coordinates of the aligned shapes if one allows for their linear dependencies by using some form of generalized inverse (this technique was used in the sampling experiments described below because it is faster). A test for differences in mean shape can be performed

using these variables and Hotelling's generalized T^2-test.

$$T^2 = \frac{n_1 n_2}{n_1 + n_2} D^2, \tag{3}$$

where

$$D^2 = (\bar{\mathbf{X}}_1 - \bar{\mathbf{X}}_2)^t \mathbf{S}^{-1} (\bar{\mathbf{X}}_1 - \bar{\mathbf{X}}_2), \tag{4}$$

$\bar{\mathbf{X}}_1$ and $\bar{\mathbf{X}}_2$ are the sample mean projections (as column vectors of length m), $\mathbf{S}$ is the pooled within-groups covariance matrix, and m is the dimensionality of shape space ($2p - 4$ for 2D data or $3p - 7$ for 3D). Note that $\mathbf{S}$ will be singular (and D^2 undefined) unless $n_1 + n_2 - 1 > m$. Under the null hypothesis, $F = (n_1 + n_2 - m - 1)/[m(n_1 + n_2 - 2)]T^2$ follows the F-distribution with m over $n_1 + n_2 - m - 1$ degrees of freedom. Note the different degrees of freedom in the denominator in comparison to Goodall's F-test. Degrees of freedom are lost here because of the need to estimate the full covariance matrix. See Jackson (1991) for a description of generalized T^2-tests, Marcus (1993) for a description of their application in morphometrics, and Dryden and Mardia (1998) for more information on their use in geometric morphometrics.

This test assumes independent samples and that the Kendall tangent space coordinates follow a multivariate normal distribution and that both samples come from populations with the same covariance matrix, Σ. The use of tangent space coordinates is based on the assumption that the shape variation is small but this does not seem to be a problem in practical applications (at least those in the biological sciences).

Bookstein shape coordinates

Bookstein (1991) shape coordinates for a triangle (also called a two-point or edge registration) can be expressed in several ways. The shape coordinates for a triangle are simply the coordinates of a free vertex, say C, after the triangle has been translated, rotated, and scaled so that one vertex, A, is at the origin (0, 0), and the other vertex, B, is at (1, 0). A convenient equation is

$$\begin{pmatrix} \nu_1 \\ \nu_2 \end{pmatrix} = \frac{1}{|B - A|^2} \begin{pmatrix} x_b - x_a & y_b - y_a \\ -(y_b - y_a) & x_b - x_a \end{pmatrix} \begin{pmatrix} x_c - x_a \\ y_c - y_a \end{pmatrix}, \tag{5}$$

where $|B - A|^2 = (x_b - x_a)^2 + (y_b - y_a)^2$, the squared length of the baseline. Note: some authors (e.g., Dryden and Mardia, 1998) subtract $\frac{1}{2}$ from ν_1 which corresponds to placing vertices A and B at $(-\frac{1}{2}, 0)$ and $(\frac{1}{2}, 0)$, respectively. If there are additional landmarks then they can each be substituted for vertex C in the above equation to yield additional shape coordinates. Bookstein shape coordinates are a special case of a stereographic projection of a point in Kendall's shape space onto a tangent plane (see above). See Figure 26 of Dryden and Mardia (1998) for an illustration.

A test for differences in mean shape in two samples can be performed using shape coordinates as variables and the generalized T^2-test as described above. Marcus (1993) shows an example.

This test assumes independent samples, that Bookstein shape coordinates have a multivariate normal distribution, and that the samples are drawn from populations with the same covariance matrix. The assumption of equality of variances can be a problem since the variance is much larger for mean shapes more distant from the reference shape (see above). The implied reference is the shape with Bookstein shape coordinates of $(\frac{1}{2}, 0)$. See Rohlf (1999b) for more information. This effect is minimized if a long baseline is used.

T^2-test using Rao and Suryawanshi (1996) shape variables

Rao and Suryawanshi (1996) proposed comparing samples of shapes using generalized distances based on the logs of the distances between all pairs of landmarks. Specifically, they used

$$\mathbf{d}^{(s)} = \mathbf{H}\mathbf{d} \tag{6}$$

as the set of $m - 1$ shape variables, where $\mathbf{d}$ is the vector of logs of the $m = (p - 1)p/2$ distances between all pairs of landmarks and $\mathbf{H}$ is an $(m - 1) \times m$ matrix of rank $m - 1$ such that $\mathbf{H1} = \mathbf{0}$ (a Helmert

matrix with the first row deleted). This equation projects a vector of log distances onto a space orthogonal to their mean (thus size-scaling by removing the log of the geometric mean distance). The matrix **H** is not unique, but its rows are orthogonal so different choices simply correspond to rotations of the space (which have no effect on distances between shapes or on the proposed statistical tests).

Rao and Suryawanshi (1996) suggested using the asymptotic relationship that $X^2 = n_1 n_2/(n_1 + n_2)D^2$ is distributed as χ^2 with $m - 1$ degrees of freedom [see Eq. (4)]. While it usually makes little difference, the F distribution was used for the T^2-test (see above) in this study since sample sizes are rather small.

This test assumes independent samples and that the log distances are normally distributed with the same covariance matrix in the two populations. The assumption of equality of covariance matrices can be a problem since the expected covariance matrix depends upon the mean shape (Rohlf, 1999a). For $p > 3$, larger sample sizes are required than for the other tests since the dimension, m, of shape space is larger than for the methods described above.

T^2-test using Rao and Suryawanshi (1998) shape variables

Rao and Suryawanshi (1998) proposed that a natural method of comparing shapes is to compare angles from a triangulation of landmarks. Since the three angles in a triangle sum to a constant (π radians), only two arbitrarily selected angles from each triangle are used as shape variables. For $p > 3$ there are many ways in which one can select the angles to be used. In this study we have selected a baseline and then used the two angles from the baseline to each of the other landmarks. This results in a set of $2(p - 2)$ angles. In a practical application, a user will probably select a logical subset of angles (see, for example, Mardia et al., 1996, and Rao and Suryawanshi, 1998) but that is difficult to take into account in a study such as the present one. Dryden and Mardia (1998) discuss the use of angles and point out that there is a loss of information and that there are pathological cases.

Samples of shapes are compared by using standard multivariate methods, e.g., T^2-tests, on these shape variables ($p - 2$ pairs of angles).

This test assumes independent samples and that the angles follow a multivariate normal distribution with the same covariance matrix in the two populations. The assumption of equality of covariance matrices can be a problem since the expected covariance matrix depends upon the mean shape (Rohlf, 1999a). While several transformations are suggested by Rao and Suryawanshi (1998) to reduce departures from normality, they were not used in this study since their examples were based on the use of angles and because departures from normality seem less important than the consequences of heterogeneous covariance matrices.

EDMA-I T statistic

Lele and Richtsmeier (1991) proposed the use of a statistic T that is computed as follows. First, compute form matrices consisting of the inter-landmark distances for each specimen. These are averaged for each sample. An improved method of averaging was suggested by Lele (1993) and used in this study. A form difference matrix is then computed as the element-wise ratios of the average interlandmark distances in the average form matrices for the two samples. Their statistic T is the ratio of the largest to the smallest of the elements of the form difference matrix. The statistical significance of T is assessed by comparing the observed value to an empirical distribution of T values from a non-parametric bootstrap procedure as described in Richtsmeier and Lele (1993).

This test assumes independent samples and the equality of the covariance matrices in the two populations being compared (Lele and Cole, 1996). The assumption of equality of covariance matrices may be a problem since the expected covariance matrix depends strongly on the mean shape (Rohlf, 1999a).

EDMA-II *Z* statistic

Lele and Cole (1995; 1996) proposed a statistic, Z, which is the maximum absolute value of the arithmetic difference between the two size-scaled average form matrices being compared. They reported that scaling by the length of a baseline resulted in higher power than scaling by centroid size. The use of centroid size scaling was not investigated in detail since there were serious problems in the type I error rates. EDMA-II was proposed by Lele and Cole (1995, 1996) as a method with higher power than EDMA-I and T^2-tests using Bookstein shape coordinates (see also electronic postings by T. Cole in 1995 and 1996 to the morphmet Internet discussion group). These conclusions were, however, based on simulations that were in error (confirmed by T. Cole, personal communication).

The statistical significance of the observed Z-value is assessed using the parametric bootstrap confidence intervals of Hall and Martin (1988). In this procedure, 100 pairs of multivariate normally distributed samples of shapes are generated with the same estimated mean and covariances as the observed data. Maximum signed differences are computed for each pair and sorted from low to high. A $100(1 - \alpha)\%$ confidence interval is given by the $100\alpha/2$ and $100(1 - \alpha/2)$ percentiles of this array (interpolating if necessary for α values such as 0.05). The null hypothesis of no difference in shape is rejected at the α level of significance if the estimated confidence interval does not contain zero.

This test assumes independent samples and normally distributed variation at each landmark, but it does not assume the equality of the covariance matrices in the two populations being compared.

RESULTS

The power surfaces shown in Figure 2 were generated using specially written software. The height of each point on the surface gives the proportion of times in which the null hypothesis of no shape difference was rejected for a pair of samples. The numerical estimates for power for particular comparisons can be computed using the tpsPower software that runs under Windows 95/98/NT. It can be downloaded from the Stony Brook morphometrics www site (http://life.bio.sunysb.edu/morph) and used to further investigate the methods described below.

Three landmarks—triangles in the plane

The methods listed in Table 1 were compared one the basis of statistical power estimated from sampling experiments. Relatively small sample sizes and large amounts of variability were used to make the differences among the methods more apparent. This also made it possible to use a coarser grid in the plots of the power surfaces (see below). This was important since simulations for some of the power surfaces required over 4 h even using a 400 MHz Pentium II computer with software specialized to triangles.

Power surfaces were constructed by holding one shape constant and then systematically varying a second shape over the range of possible triangles. This provides a useful visualization of how power for a method varies as a function of position in its shape space. See Rohlf (1999a) for a description of the shape spaces for the various methods. It is not practical to display such surfaces for the EDMA-I and EDMA-II methods, since their shape spaces are highly curved. While it would be feasible for the other methods, it would be inconvenient to have to compare surfaces that were not plotted in the same way. For these reasons, the power surfaces are all plotted with respect to Kendall tangent space. One expects power to vary from near α for shapes close to the fixed shape and then increase smoothly to 1 for shapes more distant from the fixed shape. For the first set of experiments the fixed shape was an equilateral triangle—a point at the center of Kendall tangent space for triangles in Figure 1. However, for those methods that do not distinguish shape differences due to reflection (see Table 1), the sampling experiments were restricted to the disk with radius $\sqrt{\frac{1}{2}}$ (the dashed circle in Fig. 1). For many points across the surface, power was estimated using 3,000 pairs of samples (1,000 for the EDMA methods since they required so much computer time).

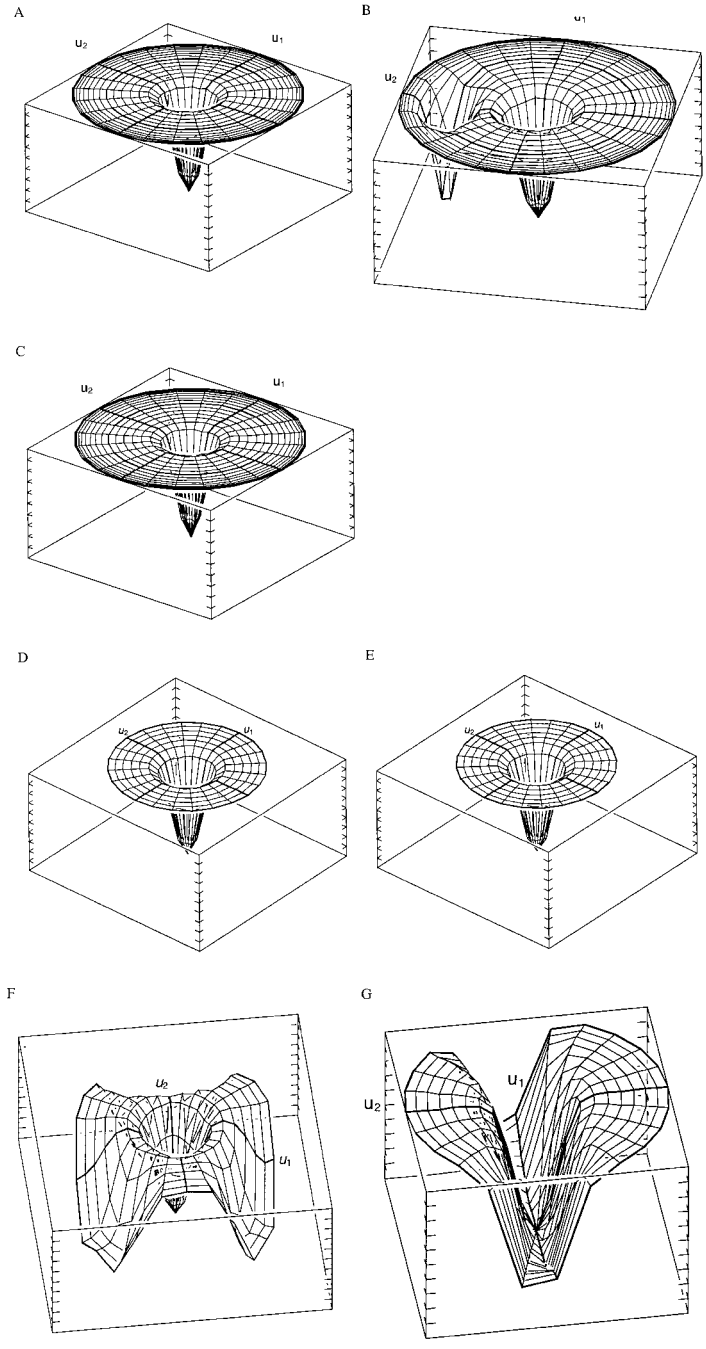

Fig. 2. Power surface for tests comparing an equilateral triangle against various alternative shapes. Surfaces plotted using Kendall tangent space coordinates. Parameters: $\sigma = 0.15$, $\alpha = 0.10$ (except for EDMA-II where $\alpha = 0.05$ was used), $n = 10$. The tic marks indicate 0.1 increments in power along the vertical axes. The other two axes range from -1 to $+1$. The surfaces in figures **(C)—(G)** are just plotted over the range $-\sqrt{1/2}$ to $\sqrt{1/2}$ because the corresponding methods ignore reflections. **(A)** KendTS, **(B)** BookSC (AB edge used as the baseline). **(C)** GoodF. **(D)** Rao-d. **(E)** Rao-a. **(F)** EDMA-I. **(G)** EDMA-II using the AB baseline for size scaling.

Before one can compare power surfaces, one must first check whether the type I error rates are sufficiently close to the intended levels. It was found that all of the methods gave satisfactory results except for EDMA-II. Using $\alpha = 0.1$ and scaling by the length of the AB baseline, type I error rates were obtained that varied from about 0.05 at the positive end of the u_2 axis (top of Fig. 1) to about 0.3 near the negative end (bot-

tom of Fig. 1). This makes it difficult to construct a power surface that can be fairly compared to other methods. The EDMA-II simulations were run with an intended type I error rate of 0.05 in order to obtain an actual type I error rate close to 0.1 to make their results comparable to those for the other methods. A value of $\alpha = 0.1$ was used for the other methods. When centroid size scaling was used for EDMA-II, the type I error rates were *much* lower than expected near the center of the shape space. A type I error rate of about $\alpha = 0.48$ would be required in order to actually achieve a type I error rate near 0.1 for shapes close to an equilateral triangle. For other kinds of triangles, the type I error rates followed a pattern similar to that described above. Because of these difficulties, power surfaces for EDMA-II were only constructed using baseline scaling.

Figure 2 shows the results for the various methods using an equilateral triangle as the mean of the fixed population. The surfaces in Figure 2A, B, and C cover the full unit disk since the corresponding methods distinguish reflections. The surface for BookSC (Fig. 2B) shows an unexpected secondary region of low power. This corresponds to comparing an equilateral triangle to isosceles triangles in which the length of the baseline is very short (compare its position relative to the u_1, u_2 axes in Fig. 1). A similar surface would be obtained if a different edge were used as a baseline—the power surface would simply be rotated by $2\pi/3$ radians.

The surfaces in Figure 2D and E are similar except that they are limited to the part of shape space that excludes reflections of the equilateral triangle. The surfaces for EDMA-I (Fig. 2F) and EDMA-II (Fig. 2G) also exclude reflections but are strikingly different and rather bizarre. This figure shows that while the EDMA-I method can distinguish an equilateral triangle from similar shapes, it has difficulty distinguishing it from more dissimilar shapes. On the other hand, the EDMA-II method (using the AB edge for size scaling) has difficulty distinguishing an equilateral triangle from triangles whose u_2 value is close to zero (points along the u_1 axis in Fig. 1). Similar results would be obtained for the EDMA-II method if a different edge was used for size scaling—the surface would simply be rotated by $2\pi/3$ radians and the direction of low power would also be rotated by $2\pi/3$ radians in Figure 1.

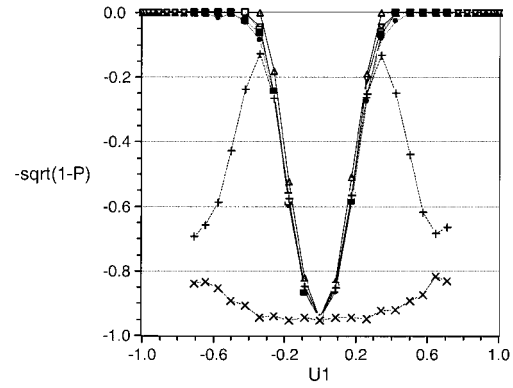

Fig. 3. Power curves as a function of position along the u_1 axis. Ordinate is $-(1 - \text{power})^{1/2}$ in order to expand the scale for larger values of power. Parameters: $\sigma = .15$, $\alpha = 0.1$. Bookstein shape coordinates = ●, Goodall F-test = △, Kendall tangent space = ▽, $\ln(d)$ = □, angles = ■, EDMA1 = +, EDMA2 = ×.

While Figure 2 gives a good overall impression, it is difficult to compare the methods at particular points. Figure 3 and 4 represent cross-sections through these surfaces along the u_1 and u_2 axes, respectively, through the surfaces in Figure 2. These figures permit more detailed comparisons among the methods. The ordinate is expressed as $-(1 - P)^{1/2}$ (where P is the estimated power) to enlarge the scale for larger values of power. Larger values correspond to higher power. In Figure 3, for $u_2 = 0$ and various values of u_1, the methods showed similar patterns of change in power except for the strange behavior of the EDMA-I and EDMA-II methods. EDMA-I shows the puzzling result of power increasing and then decreasing. Power is much higher, for example, for distinguishing a triangle corresponding to $u_1 = 0.342$ (see Fig. 1) than for a triangle corresponding to $u_1 = 0.643$ (power was 0.983 and 0.533, respectively). EDMA-II shows very low power for all values of u_1 since this cross-section goes through the middle of the trough in the power surface shown in Figure 2G. A more detailed examination of this figure shows that the GoodF method has the

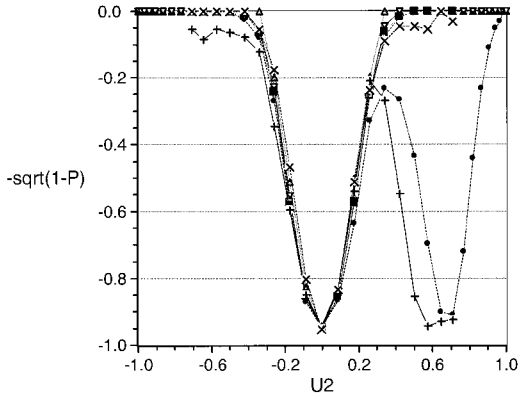

Fig. 4. Power curves as a function of position along the u_2 axis. Parameters: $\sigma = 0.15$, $\alpha = 0.1$. Ordinate is $-(1 - \text{power})^{1/2}$ in order to expand the scale for larger values of power. Parameters: $\sigma = 0.15$, $\alpha = 0.1$. Symbols: BookSC = ●, GoodF = △, KendTS = ▽, Rao-d = □, Rao-a = ■, EDMA1 = +, and EDMA2 = ×.

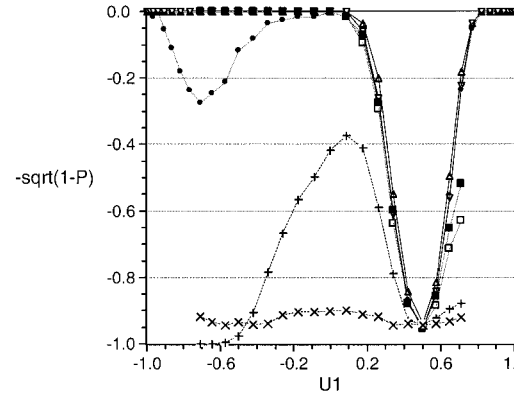

Fig. 5. Power curves as a function of position along the u_1 axis using a triangle at $u_1 = 0.574$, $u_2 = 0$ as the mean of the fixed population. Ordinate is $-(1 - \text{power})^{1/2}$ in order to expand the scale for larger values of power. Parameters: $\sigma = 0.15$, $\alpha = 0.1$. Bookstein shape coordinates = ●, Goodall F-test = △, Kendall tangent space = ▽, Ln(d) = □, angles = ■, EDMA1 = +, EDMA2 = ×.

highest power followed by KendTS, followed by the other three methods (BookSC, Rao-d, and Rao-a).

Figure 4 shown a very different pattern for $u_1 = 0$ and various values of u_2. Both EDMA-I and BookSC show low power to the right but for different reasons. Positive values of u_2 correspond to triangles where the baseline (used by BookSC but not EDMA-I) is the shortest side of the triangle. Note that power for BookSC is generally lower on the right side of the plot than on the left due to the effect of a shorter baseline. EDMA-I shows lower power to the right because that method has low power against isosceles triangles. EDMA-II shows high power—especially on the left side of this plot. Thus this method is much more sensitive to changes in shape corresponding to differences along the u_2 axis in Figure 1. This is consistent with the results posted by John T. Kent in 1995 to the morphmet Internet discussion group.

While it makes no difference for the GoodF or KendTS methods and little difference for BookSC, using an equilateral triangle for the mean of the fixed population represents the most favorable case for the other methods. To demonstrate this, the sampling experiments were repeated first using an isoceles triangle at $u_1 = 0.574$ and $u_2 = 0$ and then an oblique triangle at $u_1 = 0$ and $u_2 = 0.574$ as the means of the fixed population. In addition to the unsatisfactory behavior of the EDMA methods, Figure 5 shows that GoodF has the highest power. KendTS and BookSC have lower power. The dip in power at the left for the BookSC method corresponds to using the shortest side of a triangle as the baseline. The Rao-a and especially the Rao-d methods show much lower power. In Figure 6 the EDMA-I method again behaves very poorly. EDMA-II shows the expected pattern of power increasing as the mean shapes becomes more different but is distinctly lower. The ranking of the other methods are the same as in Figure 5. The type I error rate for the EDMA-II method is too low in these comparisons since the mean of the fixed population is near the positive end of the u_2 axis. This means that the power of the EDMA-II method would be higher if the test were adjusted to give the proper type I error rate.

Table 2 gives numerical results of sampling experiments for a test for shape difference of a pair of isosceles triangles considered by Lele and Cole (1996). In the coordinate system used for triangles in this paper, they are located at $U = (0, 0.716)$ and $(0, 0.118)$, respectively (note that they differ only with respect to their position

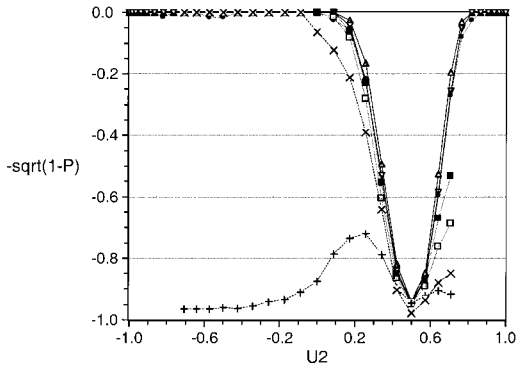

Fig. 6. Power curves as a function of position along the u_1 axis using a triangle at $u_1 = 0$, $u_2 = 0.574$ as the mean of the fixed population. Ordinate is $-(1 - \text{power})^{1/2}$ in order to expand the scale for larger values of power. Parameters: $\sigma = 0.15$, $\alpha = 0.1$. Bookstein shape coordinates = ●, Goodall F-test = △, Kendall tangent space = ▽, Ln(d) = □, angles = ■, EDMA1 = +, EDMA2 = ×.

along the u_2 axis). A value of $\alpha = 0.05$ was used for EDMA-II since, as reported above, this results in type I error rates closer to the intended level of 0.10 (the type I error rate was about 0.16 using $\alpha = 0.10$ and 0.09 using $\alpha = 0.05$). While the results are similar for most methods, Goodall's F-test has the highest power except for EDMA-II with the AB edge used for size scaling. This is the type of comparison for which EDMA-II does best—a difference in shape space parallel to the u_2 axis. Power is much lower when one of the other pairs of landmarks are used (the results for scaling by the distance between landmarks B and C are not shown since they are equivalent to those for A and C).

Lele and Cole's (1996) results were rather different from those shown here due to problems in their software. The present results agree closely with those posted by John T. Kent to the morphmet Internet discussion group in 1995 and 1996.

More than three landmarks

It is more difficult to study power systematically when there are more than three landmarks since the shape spaces are multidimensional (four dimensions even for just four landmarks in the plane or five dimensions for four landmarks in 3D). However, one expects the relationships found for triangles to generalize. We can check our understanding through selected examples.

Table 3 shows the results for a comparison of two tetrahedra used in Lele and Cole (1996). This should represent a favorable case for the inter-landmark distance based methods since the inter-landmark distances are similar. However, power was distinctly higher when using GoodF. The next highest power was for KendTS. Power for EDMA-II was low for all choices of baseline (results are only shown for three choices).

Table 4 shows the results using a pair of five landmarks 2D shapes differing only in the relative position of landmark 3 (illustrated in Rohlf, 1999a). The results are similar for most of the methods. Power for BookSC, Rao-a, and EDMA-II depended strongly on the choice of baseline. This is particularly true for the EDMA-II method. Power is very high when baseline is one of interlandmark distances involved in the shape change and very low otherwise. BookSC has high power when one of the longer distances is used for the baseline.

In both of these examples, the shape difference only involved the change in position of a single landmark (a small-scale shape difference). Table 5 shows the results for a comparison at the largest spatial scale—a uniform shape change. One shape was simply a 3×3 grid of nine equally spaced landmarks. The second shape was created by stretching the shape at 45° by a factor of 1.1 (by rotating the space and then multiplying the x coordinates by 1.1). Type I error rates were close to the intended value of 0.1 for all of the methods except for GoodF when the variance was not small. Values of σ equal to 0.1, 0.2, 0.4, 0.6, 0.8, and 1 resulted in type I error rates of 0.104, 0.14, 0.224, 0.449, 0.668, and 0.845, respectively, for the GoodF method. A value of $\sigma = 0.15$ was used in the sampling experiments. As expected, the results depend upon the choice of size scaling. Scaling by centroid size resulted in GoodF having the highest power followed by KendTS. Rao-d and EDMA-I had much lower power. Different choices of baselines resulted in similar values for power for the BookSC and Rao-a methods (Rao-a showed somewhat more variability). The results for EDMA-II depend strongly on the choice of baseline (the results for only a few of the 36 possible

TABLE 2. Power and size for tests of triangles with coordinates (−5, 0), (5, 0), (0, 10) for the first shape and (−5, 0), (5, 0), (0, 11) for the second resulting in ρ = 0.04678[1]

n	Scaling	GoodF	KendTS	BookSC	Rao-d	Rao-a	EDMA-I	EDMA-II
10	Csize or none	0.252	0.238	—	0.239	—	0.287	—
	Edge AB	—	—	0.236	—	0.232	—	0.285
	Edge AC	—	—	0.215	—	0.221	—	0.147
30	Csize or none	0.552	0.533	—	0.529	—	0.585	—
	Edge AB	—	—	0.524	—	0.496	—	0.595
	Edge AC	—	—	0.550	—	0.571	—	0.263

[1] Sampling experiments used normally distributed error with σ = 1 and tests performed at α = 0.10 (except for EDMA-II at an intended level of α = 0.05). Csize = centroid size. Empty cells correspond to invalid combinations. EDMA-I does not use size scaling.

TABLE 3. Power and size for tests of tetrahedra from lele and cole (1996)[1]

Scaling	GoodF	KendTS	Rao-d	Rao-a	EDMA-I	EDMA-II
Csize or none	0.629	0.531	0.473	—	0.401	—
Edge AB	—	—	—	0.501	—	0.288
Edge BC	—	—	—	0.540	—	0.205
Edge CD	—	—	—	0.325	—	0.223

[1] The coordinates for the first shape are (0, 0, 0), (10, 0, 0), (0, 10, 0), (0, 0, 10) and (1, 1, 1), (10, 0, 0), (0, 10, 0), (0, 0, 10) for the second resulting in ρ = 0.09722. Sampling experiments used 1,000 pairs of normally distributed samples of size 10 with σ = 1 and tests performed at α = 0.10. Csize = centroid size. Empty cells correspond to invalid combinations. EDMA-I does not use size scaling.

TABLE 4. Power and size for tests of mean shapes from figure 14 of rohlf (1999a)[1]

Scaling	GoodF	KendTS	BookSC	Rao-d	Rao-a	EDMA-I	EDMA-II
Csize or none	0.996	0.976	—	0.849	—	0.997	—
Edge 1, 2	—	—	0.952	—	0.902	—	0.110
Edge 2, 3	—	—	0.327	—	0.866	—	0.980
Edge 3, 4	—	—	0.380	—	0.918	—	0.950
Edge 4, 5	—	—	0.949	—	0.765	—	0.100

[1] The coordinates for the first shape are (0, 0), (3, 6), (4, 6), (6, 3), (11, 3) and (0, 0), (3, 6), (6, 4), (6, 3), (11, 3) for the second resulting in ρ = 0.26434. Sampling experiments used 1,000 pairs of normally distributed samples of size 10 with σ = 1 and test performed at α = 0.10. Csize = centroid size. Empty cells correspond to invalid combinations. EDMA-I does not use size scaling.

TABLE 5. Power and size for tests of a 3 × 3 grid of 9 landmarks (only selected baselines are shown)[1]

Scaling	GoodF	KendTS	BookSC	Rao-d	Rao-a	EDMA-I	EDMA-II
Csize or none	0.849	0.724	—	0.260	—	0.122	—
Edge 1, 3	—	—	0.696	—	0.730	—	0.089
Edge 1, 4	—	—	0.702	—	0.664	—	0.103
Edge 1, 5	—	—	0.693	—	0.719	—	0.335
Edge 1, 7	—	—	0.734	—	0.775	—	0.088
Edge 1, 9	—	—	0.685	—	0.751	—	0.802
Edge 2, 4	—	—	0.726	—	0.664	—	0.312
Edge 2, 6	—	—	0.678	—	0.682	—	0.293
Edge 3, 7	—	—	0.719	—	0.782	—	0.770
Edge 4, 7	—	—	0.687	—	0.643	—	0.102

[1] The coordinates for the first shape are (0, 0), (0, 3), (0, 6), (3, 0), (3, 3), (3, 6), (6, 0), (6, 3), (6, 6). The second shape is a uniform transformation of the first (created by rotating by 45° and then multiplying the x-coordinates 1.1). The two shapes differed by ρ = 0.04758. Sampling experiments used 1,000 pairs of normally distributed samples of size 20 with σ = 0.15 and tests performed at α = 0.10.

choices are shown in Table 5). Only baselines that happened to be parallel to the directions of shape change in this example (±45°) were able to detect differences in shape with a probability much larger than the intended type I error rate. Power was largest when the 1, 9 and 3, 7 baselines were used. Shorter baselines had lower power even when oriented in the direction of the known shape difference.

DISCUSSION

The simulations were performed using small sample sizes and relatively large amount of variability at each landmark in order to make the patterns of differences in

power easier to display. In most applications the variances will be smaller and the sample sizes will be somewhat larger. These both lead to the same patterns of differences among the methods—but with a smaller region of low power around the point corresponding to no true difference in shape. The relative performance of the methods is summarized below.

GoodF. This test is very well behaved. As one would expect for a reasonable statistical test, power increases uniformly as the true difference between the two populations increases. The shape and height of the surface does not depend upon the mean shape as the power surface is always centered over the mean shape. For small values of σ, this method was found to have the highest power except for those special situations in which EDMA-II has higher power (due to an optimal choice of a baseline). For large values of σ (which violate one of the assumptions of the test) the type I error rate can become very large. The general superiority of GoodF is due in part to the restrictive and somewhat unrealistic assumptions that the variance is small and homogeneous across all landmarks and that variation within and among landmarks are independent. The GoodF method was able to take advantage of the fact that those assumptions were true in the sampling experiments carried out here.

KendTS. This test is also well-behaved. While its use also assumes that variation in shape is not too large, it is much more tolerant of larger variances than is the GoodF method. Type I error rates were found to be close to their intended values even for fairly large values of σ. Power increases as the Procrustes distance between the two means increases. It also does not depend on the mean shape as long as the overall mean is used as the reference in constructing the Kendall tangent space. This method would show lower power if a reference configuration were used that was far from the observed data. This could happen if, for example, juveniles of an outgroup species were used for the reference as advocated by Zelditch et al. (1998) (but see Rohlf, 1998).

BookSC. This test performs almost as well as KendTS as long as the variation in shape is not large and the baseline is not short. For very small variation in shape this method should give results essentially identical to those of KendTS. Type I error rates were found to be close to their intended values. The main limitation of the BookSC method is that close landmarks should not be used to define a baseline since that can result in very low power. This result explains why Coward and Conathy (1996) found differences in power dependent on the choice of baseline. They also reported the unexpected result that power decreased as sample size increased for their 3D test case. That puzzling result was not found in this study.

Rao-d. The power for this test is similar to that of KendTS for comparisons between shapes near the center of shape space. Rao-d does not perform as well for shapes away from the center of shape space (i.e., shapes with unequal interlandmark distances). Type I error rates were found to be close to their intended values. The method did very poorly at detecting uniform shape changes. This may have been due in part to the fact that the distances are necessarily unequal since more than three landmarks were used in the example. The number of dimensions in the Rao-d shape space increases much more rapidly than for most of the other methods as more landmarks are considered. For a given sample size, that will lead to lower power (in comparison to the other methods) as the number of landmarks is increased.

Rao-a. This method often performs better than the Rao-d method but worse than the GoodF and KendTS methods. It performed as well as the BookSC method in detecting a uniform shape change. Type I error rates were found to be close to their intended values. A problem with the method is that the selection of angles is arbitrary. Fortunately, different choices usually gave similar results.

EDMA-I. This test performs better at distinguishing pairs of triangles close to being

equilateral, than it does for distinguishing more dissimilar triangles. This seems consistent with the fact that the EDMA-I shape space is highly curved (Rohlf, 1999a). Type I error rates were found to be close to their intended values. EDMA-I did very poorly in detecting a uniform shape change. This may have been due in part to the fact that the distances between all pairs of landmarks were necessarily unequal since there were more than three landmarks in this example. Power is expected to decrease as more landmarks are considered. Unlike our results, Coward and Conathy (1996) reported that EDMA-I had inflated type I error rates. This discrepancy might be due to the fact that they used a more complex error structure or that they did not use Lele's (1993) improved method for estimating the mean form. They also reported the unexpected result that power decreased as sample sizes increased. That relationship was not found in the present study.

EDMA-II. The performance results for this method are difficult to summarize. One complication is that actual type I error rates depended on the shapes being compared. That needs to be taken into consideration when using this method in a practical application. Certain shape differences are difficult to detect using the EDMA-II method. John T. Kent correctly predicted this in his 1995 posting to the morphmet Internet discussion group. For the same pairs of shapes power can be very high using certain baselines and very low for others. It is clear that the strategy for selecting the best baseline to use for "size scaling" is very different from that used for the BookSC method. For the BookSC method one selects a well-separated pair of landmarks as a measure of the overall size of the organism. For EDMA-II the best baseline is the pair of landmarks whose separation is most influenced by the shape change one wishes to detect. While (Lele and Cole, 1996) argue strongly against it, there would be a strong temptation in a practical application to try a variety of baselines to find the best one or to at least select a baseline in a region where one suspects there are shape differences. This is because the differences in power are so large. Such a practice would, of course, would require one to adjust the levels of significance obtained (using, for example, Bonferroni methods, Sokal and Rohlf, 1995). Alternatively, this property might lead to a useful procedure for detecting influential pairs of landmarks.

In general, the sampling experiments show that GoodF has the highest power followed by KendTS. BookSC does nearly as well except when the length of its baseline is short (a situation that users avoid). The Rao-d and Rao-a methods do better when there are few landmarks and the inter-landmark distances or angles are nearly equal. However, these methods never have the highest power. They also do not offer any particular computational advantages or provide efficient methods for displaying complex shape differences. The EDMA-I method for triangular shapes does well only in a few situations. It does worse as one increases the number of landmarks and the interlandmark distances become more unequal. The EDMA-II method does best only if one selects for the baseline the pair of landmarks most influenced by the shape difference one is trying to detect. The type I error rates for this method are also often not near their intended values so one should not apply the test without first checking type I error rates for the types of shapes being compared. These problems make the EDMA-II method much less desirable as a general tool for testing for shape differences—especially in exploratory studies. The results of the sampling experiments reported here are not surprising (except for seeing just how bizarre the power surfaces are for the EDMA methods). The GoodF method and its KendTS and BookSC approximations are based on rigorous sets of axioms and theory for shape built upon the work of Kendall (1981, 1984). One expects the GoodF method to have the highest power if its rather strict assumptions are met. Otherwise one expects the KendTS method to have the highest power. The BookSC method should give very similar results when shape variation is very small and the baseline is not short.

It is clear that additional work needs to be done to investigate robustness of these

methods with respect to such complications as unequal variances, outliers, correlations among landmarks, etc. While Coward and Conathy (1996) considered such models, only a single 2D and 3D cases were investigated so that it is not possible to draw any general conclusions. For such models the relative ranking of the various methods could be rather different than for the ideal case studied here. The ranking of the GoodF method is quite likely to change since it requires the most stringent assumptions. Sampling experiments are also needed to investigate the usefulness of bootstrap and permutation test methods for the Goodall's F and the T^2 statistics for these more complex (and hence more realistic) models. It seems likely that these procedures will do very well.

ACKNOWLEDGMENTS

The critical comments and many suggestions by Fred L. Bookstein are gratefully acknowledged. The critical reading of the manuscript by Dean Adams was very helpful. The comments by Dennis Slice and Waleed AlGharaibeh were also appreciated. This paper is contribution no. 1052 from the Graduate Studies in Ecology and Evolution, State University of New York at Stony Brook.

LITERATURE CITED

Bookstein FL. 1991. Morphometric tools for landmark data: Geometry and biology. New York: Cambridge University Press.

Bookstein FL. 1996. A standard formula for the uniform shape component in landmark data. In: Marcus LF, Corti M, Loy A, Naylor GJP, Slice DE, editors, Advances in morphometrics, Vol. 284, New York: Plenum. p 153–168.

Bookstein FL. 1998. A hundred years of morphometrics. Acta Zool 44:7–59.

Coward WM, Conathy DM. 1996. A Monte Carlo study of the inferential properties of three methods of shape comparison. Am J Phys Anthropol 99:369–377.

Dryden IL, Mardia KV. 1992. Size and shape analysis of landmark data. Biometrika 79:57–68.

Dryden IL, Mardia KV. 1998. Statistical shape analysis. New York: John Wiley & Sons.

Goodall CR. 1991. Procrustes methods in the statistical analysis of shape (with discussion and rejoinder). J Roy Stat Soc Ser B 53:285–339.

Hall P, Martin MA. 1988. On the bootstrap and two sample problems. Austral J Stat 30A:179–192.

Jackson JE. 1991. A user's guide to principal components. New York: John Wiley & Sons.

Kendall DG. 1981. The statistics of shape. In: Barnett V, editor. Interpreting multivariate data. New York: John Wiley & Sons. p 75–80.

Kendall DG. 1984. Shape-manifolds, Procrustean metrics and complex projective spaces. Bull Lond Math Soc 16:81–121.

Kendall DG. 1985. Exact distributions for shapes of random triangles in convex sets. Adv Appl Prob 17: 308–329.

Kent JT. 1994. The complex Bingham distribution and shape analysis. J Roy Stat Soc Ser B 56:285–299.

Lele S. 1993. Euclidean distance matrix analysis: estimation of mean form and form difference. Math Geol 25:573–602.

Lele S, Cole TM III. 1995. Euclidean distance matrix analysis: a statistical review. Current issues in statistical shape analysis, Vol. 3. Leeds: University of Leeds. p 49–53.

Lele S, Cole TM III. 1996. A new test for shape differences when variance-covariance matrices are unequal. J Hum Evol 31:193–212.

Lele S, Richtsmeier JT. 1991. Euclidean distance matrix analysis: a coordinate free approach for comparing biological shapes using landmark data. Am J Phys Anthropol 86:415–427.

Marcus LF. 1993. Some aspects of multivariate statistics for morphometrics. In: Marcus LF, Bello E, Garcia-Valdecasas A, editors. Contributions to morphometrics, Vol. 8. Madrid: Monographias del Museo Nacional de Ciencias Naturales. p 95–130.

Marcus LF, Corti M, Loy A, Naylor GJP, Slice DE. 1996. Advances in morphometrics. New York: Plenum.

Mardia KV, Coombes A, Kirkbride J, Linney A, Bowie JL. 1996. On statistical problems with face identification from photographs. J Appl Stat 23:655–675.

Rao CR, Suryawanshi S. 1996. Statistical analysis of shape of objects based on landmark data. Proc Natl Acad Sci 93:12132–12136.

Rao CR, Suryawanshi S. 1998. Statistical analysis of shape through triangulation of landmarks: a study of sexual dimorphism in hominids. Proc Natl Acad Sci 95:4121–4125.

Reyment RA, Blackith RE, Campbell NA. 1984. Multivariate morphometrics, 2nd ed. London: Academic Press.

Richtsmeier JT, Cole TM III, Valeri CJ, Lele S. 1998. Preoperative morphology and development in sagittal synostosis. J Craniofacial Genet Dev Biol 18:64–78.

Richtsmeier JT, Lele S. 1993. A coordinate free approach to the analysis of growth patterns: models and theoretical considerations. Biol Rev 68:381–411.

Rohlf FJ. 1995. Multivariate Analysis of shape using partial-warp scores. In: Mardia KV, Gill CA, editors. Proceedings in Current issues in statistical shape analysis. Leeds: University of Leeds. p 154–158.

Rohlf FJ. 1998. On applications of geometric morphometrics to studies of ontogeny and phylogeny. Syst Biol 47:147–158.

Rohlf FJ. 1999a. On the use of shape spaces to compare morphometric methods. Hystrix, in press.

Rohlf FJ. 1999b. Shape statistics: Procrustes superimpositions and tangent spaces. J Classif, 16:197–225.

Rohlf FJ. 1999c. tpsTri. Department of Ecology and Evolution, State University of New York at Stony Brook, Stony Brook, NY.

Rohlf FJ, Marcus LF. 1993. A revolution in morphometrics. Trends Ecol Evol 8:129–132.

Small CG. 1996. The statistical theory of shape. New York: Springer.

Sokal RR, Rohlf FJ. 1995. Biometry. The principals and practice of statistics in biological research, 3rd ed. San Francisco: W. H. Freeman.

Zelditch ML, Fink WL, Swiderski DL, Lundrigan BL. 1998. On applications of geometric morphometrics to studies of ontogeny and phylogeny: a reply to Rohlf. Syst Biol, 47:159–167.